PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| (51) International Patent Classification 5 : C07K 3/04 | A1 | (11) International Publication Number: WO 93/20098 |
| | | (43) International Publication Date: 14 October 1993 (14.10.93) |

(21) International Application Number: PCT/US93/02846

(22) International Filing Date: 26 March 1993 (26.03.93)

(30) Priority data:
07/865,368   7 April 1992 (07.04.92)   US

(71) Applicant: THE SCRIPPS RESEARCH INSTITUTE [US/US]; 10666 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors: KENT, Stephen, B., H. ; 2766 Costebelle Drive, La Jolla, Ca 92037 (US). SCHNOLZER-RACKWITZ, Martina ; 8533-C Villa La Jolla Drive, La Jolla, CA 92037 (US).

(74) Agents: FITTING, Thomas et al.; The Scripps Research Institute, 10666 North Torrey Pines Road, TPC-8, La Jolla, CA 92037 (US).

(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: PROCESS FOR PREPARING MODIFIED PROTEINS

(57) Abstract

Process for the preparation of modified proteins comprising the coupling of a first peptide segment having a haloacyl group at the N-terminus thereof with a second peptide segment having a carbonylthiol group at the C-terminus thereof are disclosed. Novel modified proteins produced by the process are also disclosed.

Serial No.: 09/144,838
Inventor: Siani et al.
Attny Docket No. GRFN-020/01US

F3

-1-

PROCESS FOR PREPARING MODIFIED PROTEINS

BACKGROUND OF THE INVENTION

5   ## FIELD OF THE INVENTION

This invention is in the area of modified biomolecules
and methods of making such modified biomolecules.
More particularly, this invention relates to protein
10   engineering by chemical means to produce modified
proteins where one or more peptide bonds are
substituted by non-peptide linkage and one or more
encoded amino acids may be replaced by unnatural amino
acids or amino acid analogs or any other non-coded
15   structure.

RELATED ART

Numerous attempts have been made to develop a
20   successful methodology for synthesizing modified
biomolecules such as proteins, glycoproteins,
nucleotides, polysaccharides, and other biopolymers.
Such modified biomolecules are invaluable for study of
structure-activity relationships of native
25   biomolecules and there is a growing number of
commercial applications of these molecules for
diagnostic or therapeutic purposes.

Structural modification of proteins and peptides,
30   normally referred to as "protein engineering" involves
the rationally designed alteration of structure with
the aim of understanding protein structure and
function and of creating a protein with new desirable
properties. In the past, this has been principally
35   carried out by site-directed mutagenesis or other

- 2 -

techniques involving genetic manipulation. The major
drawbacks of these prior art approaches are that amino
acids replacing native amino acids are those that must
be coded genetically. As a result, other structural
5   variants such as unnatural amino acids or amino acid
analogs cannot be introduced in the protein backbone.
However, recent findings (Ellman, et al., Science,
255:197, 1992; Noren, et al., Science, 24:182, 1989)
would allow unnatural amino acids or amino acid
10   analogs to be incorporated into proteins in a site-
specific manner. In this approach, a codon encoding
an amino acid to be replaced is substituted by the
nonsense codon TAG by means of oligonucleotide-
directed mutagenesis. A suppressor tRNA directed
15   against this codon is then chemically aminoacylated
with the desired unnatural amino acid. Addition of
the amino acylated tRNA to an in vitro protein
synthesizing system programmed with the mutagenized
DNA directs the insertion of the prescribed amino acid
20   into the protein at the target site. Taking the
enzyme T4 lysozyme, the above authors, incorporated a
wide variety of amino acid analogs into the enzyme at
alanine 82 position with a few exceptions, for
example, of D-alanine not being incorporated.
25

While Schultz's approach partially solves problems
posed by biosynthetic protein engineering, it does not
allow the alteration of the protein backbone at more
than one target site to incorporate two or more
30   different non-coded structural units. Also, by the
very nature of the system, i.e., the fact that it
relies upon a living system to produce the engineered
protein, many substitutions or alterations, such as
those which would result in a lethal mutation, cannot

- 3 -

be done.   The chemical synth sis would ov rcome the
shortcomings left by the Schultz techniques.
(reviewed by R. E. Offord, *Protein Eng.*, $\underline{1}$:151, 1987).
However, chemical synthesis is fraught with many
5   difficulties such as the need of protection of
unwanted reactive groups.

Overall, there is a definite need for a simple and
efficient method for making a modified protein which
10   posses desired properties.   The present invention
addresses such need and provides novel modified
proteins.

SUMMARY OF THE INVENTION
15

This invention provides new and useful modified
biomolecules.   It also provides a new process for
producing such modified biomolecules.   In general, the
modified biomolecules of this invention comprise two
20   molecular segments, each selected from peptides,
pseudopeptides, or non-peptide linear molecules linked
through a non-amido linkage to form a peptide or
pseudopeptide backbone, wherein one or the segment
contains at least one non-coded structural unit and
25   the non-coded structural unit does not form a part of
the non-amido linkage.   The chemical bonding of the
two segments is by means of terminal reactive groups
on one segment which react with reactive groups of the
other segment molecule.
30

The process of this invention provides a directed
ligation of the two molecular segments to create a
desired bond at the ligation point(s) and comprises
the steps of:

- 4 -

     a.   providing a first segm nt having at least
             one non-coded structural unit and attaching
             a first chemoselective synthon to the first
             segment at the terminal position thereof;

5

     b.   providing a second segment optionally
             containing at least one non-coded structural
             unit, and second chemoselective synthon at
             the terminal position thereof, the second

10
             chemoselective synthon being complementary
             to the first chemoselective synthon of the
             first segment; and

     c.   ligating the first segment and the second
15
             segment, whereby the first synthon of the
             first segment and the second synthon of the
             second segment forms a non-peptide linkage,
             wherein the first segment and the second
             segment are each selected from peptides,
20
             pseudopeptides, or non-peptide linear
             molecules, provided that both segments are
             not non-peptide linear molecules at the same
             time.

25   The above sequence a-c can be repeated by using a
     first modified biomolecule as the first segment to
     which a second segment or a second biomolecule is
     ligated. The present process also may include the
     step of ligating additional segments with the first
30   and second segments which have been provided with
     additional terminal synthons that are compatible with
     the first and second synthons and chemoselective to
     synthons of the additional segments.

- 5 -

The pres nt inv ntion is th refore applicable in the
chemical synthesis of various protein conjugates, such
as proteins with reporter molecules, radionuclides,
cytotoxic agents, nucleotides, antibodies, and non-
5   protein micromolecules.

Preferably, the process of this invention involves a
series of steps comprising:

10   a.   sequentially coupling selected amino acids
                 or amino acid analogs to a terminal amino
                 acid or amino acid analog bound to a first
                 resin support to form a first peptide
                 segment-resin, the first peptide segment
15   having about two to about one hundred amino
                 acid residues;

            b.   covalently attaching a haloacyl moiety to
                 the N-terminus of the first peptide segment-
20   resin to form a haloacylpeptide segment
                 bound to the first resin support;

            c.   cleaving the haloacylpeptide peptide segment
                 from the first resin support;
25
            d.   sequentially coupling selected amino acids
                 or amino acid analogs to a terminal amino
                 acid or amino acid analog bound to a second
                 resin support through a sulfur or selenium-
30   containing bond to form a second peptide
                 segment-resin, the second peptide segment
                 having about two to about one hundred amino
                 acid residues;

- 6 -

    e. cleaving the second peptide s gment-resin to form a second peptide segment having a thiol- or selenol-containing group at the C-terminus thereof; and

    f. coupling the haloacylpeptide peptide segment and the second peptide segment to form a modified polypeptide.

The order of the sequence of steps a-b-c-d-e- is not critical to this invention. The sequence of steps a-b-c and the sequence of steps d-e may be conducted successively or separately. The entire sequence can be repeated in a chain-reaction manner.

Optionally, any reactive groups such as thiol that may be present in the peptide segments can be protected prior to step (f) and deprotected after step (f) is complete.

This invention, in its broadest sense, encompasses a biologically active protein comprising two molecular segments, each selected from peptides, pseudopeptides, or non-peptide linear molecules linked through a non-amido linkage to form a peptide or pseudopeptide backbone, wherein one of the segments contains at least one non-coded structural unit and the non-coded structural unit does not form a part of the non-amido linkage, provided that both segments are not a non-peptide linear molecule at the same time.
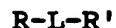
This invention further provides a modified protein represented by the formula:

- 7 -

R-L-R'

wherein R and R' are the same or different and are
each a residue of a peptide or pseudopeptide; and L
5   represents a thiol ester or selenol ester linkage.
Preferably, both R and R' comprise from about two to
about one hundred amino acid residues.

The above objects and features of the invention will
10   become more fully apparent from the description of the
preferred embodiments in conjunction with the
accompanying figures.

BRIEF DESCRIPTION OF THE FIGURES
15

FIGURE 1 shows a synthetic strategy for the total
chemical synthesis of HIV-1 PR analogs in accordance
with this invention where "A" schematically represents
the coupling of the N-terminal segment HIV-1 PR (1-50,
20   Gly^{51} SH) and the C-terminal segment bromoacetyl (53-
99) HIV-1 PR, and "B" schematically represents the
coupling of the N-terminal segment HIV-1 PR (1-50,
Cys^{51}amide) and the same C-terminal segment.

25   FIGURE 2 shows an elution profile of aliquots taken at
t = 0, 45 min., 3 h, and 48 h from the ligation
mixtures containing HIV-1 PR (1-50, Gly^{51}SH) and
bromoacetyl (53-99) HIV-1 PR by the practice this
invention using reverse phase HPLC (absorbance 214nm).
30

FIGURE 3A shows the ion spray mass spectrum of the
purified ($[NHCH_2COSCH_2CO]^{51-52}Aba^{67,95}$) HIV-1 PR where the
labeled peaks represent a single molecular species
differing in the number of excess protons.

- 8 -

FIGURE 3B shows the deconvoluted mass spectrum of the purified ([NHCH$_2$COSCH$_2$CO]$^{51-52}$ Aba ^{62,95}) HIV-1 PR where the peak of the molecular weight of the enzyme is located at 10, 769 dalton.

FIGURE 4 shows an elution profile of aliquots taken at t = 0, 30 min, and 120 min from the ligation mixture containing HIV-1 PR(1-50, Cys amide) and bromoacetyl (53-99)] HIV-1 PR.

FIGURE 5 shows an elution profile of the hexapeptide Ac-Thr-Ile-Nle-Nle-Gln-Arg; NH$_2$ before and after treatment with the ligation mixture taken at t = 3h as shown in FIGURE 2 using reverse phase HPLC with absorbance monitored at 214 nm, where the upper panel shows peaks before the treatment and the lower panel, peaks after the treatment, respectively.

FIGURE 6 shows a fluorogenic assay of aliquots of the ligation mixture taken at the times indicated as shown in FIGURE 2 where data points illustrating fluorescence units were read from continuous chart recorder tracings.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

This invention is based on a conceptually novel approach to synthesizing large biomolecules such as proteins by a convergent type synthesis. Thus, when applied to the synthesis of modified proteins, this invention involves the coupling of at least two peptide segments to create a linkage which can be a non-amido bond. The peptide segments may be synthesized using solid phase peptide synthesis,

- 9 -

solution phase synthesis, or by other techniques known
in the art including combinations of the foregoing
methods.

5   The present process involves the steps of:

     a.   providing a first segment having at least
           one non-coded structural unit, and a first
           chemoselective synthon at the terminal
10   position thereof;

     b.   providing a second segment optionally
           containing at least one non-coded structural
           unit, and a second chemoselective synthon at
15   the terminal position thereof, the second
           chemoselective synthon being complementary
           to the first chemoselective synthon of the
           first segment; and

20   c.   ligating the first segment and the second
           segment, whereby the first synthon of the
           first segment and the second synthon of the
           second segment forms a non-peptide linkage,
           wherein the first segment and the second
25   segment are each selected from peptides,
           pseudopeptides, or non-peptide linear
           molecules, provided that both segments are
           not non-peptide linear molecules at the same
           time.
30

Preferably, both of the first and second segments are
peptides or pseudopeptides having from about two to
about one hundred amino acid residues.   More

- 10 -

preferably, the segments constitute from about forty
to sixty amino acid residues.

Preferably, the non-peptide linkage formed is
5   represented by one of the following linking moieties:

$-CH_2-S-$, $-CH_2-Se-$, $-CO-S-$, $-CO-Se-$, $-CH_2-NH-$,
or $-CH(OH)-CH_2-$.

10   The present process allows at least two different non-
coded structural units to be incorporated to the
segments as well as allowing the same non-coded
structural units to be incorporated at two sites.

15   In view of the broad and varied class of modified
proteins which may be bound by this invention, all
chemically modified proteins having such
characteristics as defined above are deemed to be
within the scope of this invention. However, for
20   purposes of illustrative clarity and ease of
comprehension, this invention will be described herein
in more detail utilizing a modified HIV-1 protease as
an embodiment. It should be understood that the use
of this particular enzyme for descriptive purposes
25   shall not restrict nor limit the use of other proteins
or peptides.

As employed herein, the term "modified protein" is
intended to include oligopeptides,
30   oligopseudopeptides, polypeptides, pseudopolypeptides,
and modified native proteins-synthetic or otherwise
derived. The term "pseudopeptide" means a peptide
where one or more peptide bonds are replaced by non-
amido bonds such as ester or one or more amino acids

- 11 -

are replaced by amino acid analogs. The term
"peptides" refers not only to those comprised of all
natural amino acids, but also to those which contain
unnatural amino acids or other non-coded structural
5   units. The terms "peptides", when used alone, include
pseudopeptides. The "modified proteins" have utility
in many biomedical applications because of increased
stability toward *in vivo* degradation, superior
pharmacokinetics, and enhanced or diminished
10   immunogenecity compared to their native counterparts.

HIV-1 protease (HIV-1 PR) is a virally-encoded enzyme
which cuts polypeptide chains with high specificity
and which is essential for the replication of active
15   virions (N.E. Kohl, *et al., Proc. Natl. Acad. Sci.,*
U.S.A., 85:4686, 1988). The 21,500 dalton HIV-PR
molecule is made up of two identical 99 amino acid
polypeptide chains.

20   Comparison of the crystal structures of the empty (A.
Wlodawer, *et al., Science,* 245:616, 1989) and
inhibitor-bound (for example, M. Miller, *et al.,*
Science, 246:1149, 1989) enzyme revealed that on
binding a substrate-derived inhibitor the HIV-1
25   molecule undergoes significant conformational changes
which are particularly pronounced in two exterior,
functionally-important "flap" regions. From these
crystallography studies it appears that peptide bonds
in the flap regions of the HIV-1 PR polypeptide
30   backbone are involved in the formation of ß-sheet/ß-
turn structure, in the interaction which occurs
between the two subunits of the active dimer at the
tip of each flap in the enzyme-inhibitor (substrate),
complex, and in hydrogen bonding interactions with

- 12 -

bound peptide inhibitors (and, presumably, substrates). Mutagenesis studies carried out with recombinant HIV-1 PR showed that the flap region is highly sensitive to changes in the amino acid sequence
5   (D.D. Loeb, et al., Nature, 340:397, 1989).

These observations make the flap region especially interesting as a target for protein backbone modifications to investigate the role of peptide bond
10 interactions in HIV-1 protease activity.

In the case of HIV-1 PR where modifications of the flap region are desired, this invention allows pseudo-peptide bonds to be introduced into the region. The
15 Gly^{51}-Gly^{52} bond of HIV-1 PR is particularly preferred for bond manipulation for two reasons. First, glycine is the only achiral amino acid and therefore there is no concern about loss of optical purity; and second, the Gly^{51}-Gly^{52} bond is located near the middle of the
20 99 amino acid HIV-1 PR monomer polypeptide chain. This means that the two peptide segments which are to be coupled are each about 50 residues in length.

The solid phase peptide synthesis method is generally
25 described in the following references: Merrifield, J. Am. Chem. Soc., 888:2149, 1963; Barany and Merrifield, In the Peptides, E. Gross and J. Meinenhofer, Eds., Academic Press, New York, 3:285 (1980); S.B.H. Kent, Annu. Rev. Biochem., 57:957 (1988). By the solid
30 phase peptide synthesis method, a peptide of a desired length and sequence can be produced through the stepwise addition of amino acids to a growing peptide chain which is covalently bound to a solid resin

- 13 -

particle. Automat d synthesis may be employ d in this
method.

Accordingly, these embodiments can be accomplished by
5   the steps of:

1.   sequentially coupling amino acids to a terminal
      amino acid bound to a resin support to form a
      peptide segment-resin; and

10

2.   cleaving the peptide segment from the resin
      support.

In the preferred application of this method, the C-
15   terminal end of the growing peptide chain is
     covalently bound to a -OCH_2 PAM resin and amino acids
     having protected α-amino groups are added in the
     stepwise manner indicated above. A preferred α-amino
     protecting group is the tert-butyloxycarbonyl (BOC)
20   group, which is stable to the condensation conditions
     and yet is readily removable without destruction of
     the peptide bonds or racemization of chiral centers in
     the peptide chain. At the end of the procedure the
     product peptide is cleaved from the resin, and any
25   remaining protecting groups are removed by treatment
     under acidic conditions such as, for example, with a
     mixture of hydrobromic acid and trifluoroacetic acid,
     with trifluoromethane sulfonic acid or with liquified
     hydrogen fluoride.

30

In the case of HIV-1 protease, the C-terminal segment
comprises the following amino acid sequence: F^{53}
$IKVRQYD^{60}$ $QIPVEICGHK^{70}$
$AIGTVLVGPT^{80}PVNIIGRNLL^{90}TQIGCTLNF^{99}$. The N-terminal

- 14 -

segment comprises the following amino acid sequence: $P^1QITLWQRPL^{10}$ $VTIRIGGQLK^{20}$ $EALLDTGADD^{30}$ $TVLEEMNLPG^{42}$ $KWKPKMIGGI^{50}G^{51}$. The synthesis of these segments normally require about 15 hours at a standard cycle speed.

If desired, any amino acids in the above sequences can be replaced by amino acid analogs or amino acid mimetic compounds known in the art. Suitable amino acid substitutes include ß-alanine, L-α-aminoisobutyric acid, L-α-amino-n-butyric acid (Aba), 3.4-dehydroproline, homoarginine, homocysteine, homoproline, homoserine, 3-mercaptopropionic acid, norleucine (Nle), penicillamine, pyroglutamic acid and sarcosine.

L-amino acids and D-amino acids may be used in this invention, particularly, D-amino acids are useful for the formation of a "reversible peptide sequence".

The "reversed" or "retro" peptide sequence as discussed above refers to that part of an overall sequence of covalently-bonded amino acid residues (or analogs or mimetics thereof) wherein the normal carboxyl-to-amino direction of peptide bond formation in the amino acid backbone has been reversed such that, reading in the conventional left-to-right direction, the amino portion of the peptide bond precedes (rather than follows) the carbonyl portion (see, generally, Goodman, M. and M. Chorev, *Accounts of Chem. Res.*, 12:423, 1979).

The "reversed" peptides are within the meaning of the "peptides" used throughout the specification. D-amino

acids, amino acid analogs and amino acid mimetic
compounds are collectively referred to herein as "non-
coded amino acids." Although the solid phase peptide
synthesis theoretically enables one skilled in the art

5   to prepare a peptide backbone of any length, the
efficiency of coupling amino acids (addition of an
amino acid in successive cycles) would necessarily
limit the use of this technique when a peptide to be
synthesized has greater than 150 residues.

10   Preferably, for practical and economic reasons, the
process of this invention employs from about two to
about one hundred cycles.

Accordingly, the modified proteins obtained by the

15   process of this invention have the general structure
set forth in formula (1) and R and R' are as
previously defined. If a larger peptide segment
(containing greater than one hundred amino acid
residues) is desired, such peptide segment may be

20   available from naturally occurring proteins (native
proteins) by enzymic or chemical degradation.
Alternatively, the present process can be repeated
with different ligation modes for building such large
peptide segments.

25
An important feature of this invention is the bond
formation linking two peptide segments (namely R and
R'). A variety of bond forming reactions can be used
to covalently link the two peptide segments.

30   Representative combinations of such groups are
carbonylthiol with halo to form a thiol ester bond
between the two segments, carbony selenol with melo to
form a seleno ester, or thiol with thiol to form a
disulfide bond, thiol with halo to form a thioether

bond, selenol with halo to form a selenol ether, amino
with isothiocyanate to form a thiourea bond, amino
with aldehyde to form a imine bond which can be
reduced to a carbon-nitrogen bond, thiol with
5   maleimide to form a thioether bond, and gem-diol with
boron to form oxygen-boron bonds, and hydroxyl with
carboxyl to form an ester bond. Among these,
preferred linkages are those already enumerated by way
of their structure. A preferred linkage is a sulfur-
10   containing or selenium-containing linkage which does
not readily hydrolyze *in vivo* or which is more liable
than an amido bond *in vivo*.

The most preferred linkage L is a thiol ester linkage.
15   This linkage can be accomplished by first attaching a
facile leaving group to a first peptide segment and by
attaching carbonylthiol functionality to a second
peptide segment followed by nucleophilic substitution
where the sulfur nucleophile attacks the leaving
20   group. Preferably, a haloacyl (e.g., haloacetyl) such
as iodo, chloro, or bromoacetyl is attached to the N-
terminus of the first peptide segment. The suitable
haloacyl group may be straight or branched
(substituted by alkyl). This step is conveniently
25   carried out while the first peptide segment is still
bound to a resin support.

To introduce a bromoacetyl group to the first peptide
segment, suitable activated forms of bromoacetic acid
30   may be employed in this invention. The particularly
preferred agent for this purpose is bromoacetic
anhydride.

- 17 -

When the aforementioned C-terminal segment of HIV-1 PR
is derivatiz d with the N-bromoacetylating agent, the
unprotected N-terminus of the protected peptide
segment on resin is condensed with bromoacetic
5   anhydride to produce bromoacetyl (53-99) HIV-1 PR.
Deprotection and release of the product peptide
segment from the resin support can be accomplished by
standard conditions (e.g., treatment with anhydrous HF
containing 10% p-cresol at 0°C for several hours).
10   The product peptide segment is precipitated, dried by
lyophilization, and purified, if desired, by reverse
phase HPLC, according to standard techniques known in
the art.

15   To introduce carbonylthiol functionality to the second
peptide segment, its terminal carboxylic acid moiety
may be converted to a carbonylthiol group.

When the aforementioned N-terminal segment of HIV-1 PR
20   is to be derivatized accordingly, 4-[α-(Boc-Gly-S)
benzyl] phenoxyacetamidomethyl-resin is used as the
resin support. If any amino acid or amino acid analog
other than glycine is desired, that amino acid can be
loaded on an aminomethyl-resin in the form of 4-[α-
25   (Boc-X-S) benzyl] phenoxyacetic acid wherein X
represents the amino acid.

The N-terminal derivatized peptide segment can thus
readily be prepared by the stepwise solid phase
30   synthesis. The second product peptide segment is
cleaved from the resin support, deprotected, and
isolated in the manner described above.

- 18 -

The thiol ester linkage can be generated by coupling
the two segments under normal ligation conditions.
The formation of the thiol ester linkage is highly
chemoselective and thus compatible with most reactive
5   groups that may be present in the molecules.

A typical ligation reaction is carried out by mixing
the unprotected N- and C- terminal segments in 6M
guanidine hydrochloride buffer at about pH 3-6. In
10   this buffer, the solubility of the unprotected peptide
segments is very high, thus eliminating the major
drawback of the prior art techniques having to use
protected peptide fragments despite their limited
solubility in ligation media. Other denaturants such
15   as urea, detergents, and sodium dodecyl sulfate can be
used as the ligation buffer.

The coupling of HIV-1 PR (1-50, Gly-SH) with
bromoacetyl (53-99) HIV-1 PR is complete in several
20   hours. The product peptide [(NH $CH_2COSCH_2CO)^{51-52}$] HIV-
1 PR can be isolated, purified and characterized by
standard techniques.

When one of the peptide segments has any reactive
25   groups which may interfere with the thiol ester
formation, those groups may be protected by suitable
protecting groups known in the art or the amino acids
bearing such groups can be substituted by other amino
acids or amino acid analogs which are reaction inert
30   and yet do not adversely affect the biological
activity of a product protein.

For example, in the Examples, the C-terminal segment
of HIV-1 PR has two cysteine residues at positions 67

- 19 -

and 95.   These cysteine positions have been shown to
be replaceable by L-α-amino-n-butyric acid (Aba)
without causing the loss of the enzymic activity of
the native protease.   In one embodiment of this
5   invention, bromoacetyl [(53-99)Aba^{67,95}] HIV-1 PR is
used.   Alternatively, to block a thiol group of the
cysteine residues before the coupling reaction, the
use of protecting groups that are compatible with
ligation conditions are preferred.   However, these
10   precautions may not be necessary since, in present
experience, the thiol ester group of the N-terminal
segment attacks a bromoacetyl group even at low pH
conditions where a thiol side chain of a cysteine
residue is unreactive.
15

In another embodiment of the invention, the linkage L
can be a thioether bond.   This linkage can be
accomplished in substantially the same manner as that
described for the formation of the thiol ester
20   linkage, except that thiol functionality is attached
to a second peptide segment.   This is conveniently
accomplished by utilizing a thiol side chain of a
cysteine residue.   Thus, cysteine can be employed as
the C-terminal amino acid of the second peptide
25   segment (N-terminal segment).   HIV-1 PR(1-50, Cys^{51}
amide) described herein represents one example of such
derivatized second peptide segment where the
cysteine's carboxylic acid is blocked as amide.   When
this peptide segment is subjected to the ligation
30   conditions described earlier in the presence of
bromoacetyl (53-99) HIV-1 PR, the coupling of the two
peptide segments takes place rapidly to afford
{[NHCH(CONH$_2$)CH$_2$SCH$_2$CO]$^{51-52}$}HIV-1 PR.   The ligated
product protein can be isolated, purified, and

- 20 -

characterized by standard techniques. In a further
embodiment of the invention, the linkage can be a
selenium-containing bond such as selenol ester and
seleno ether. In like manner, a selenocysteine can be
5   used in place of cysteine.

The process of this invention is applicable to the
synthesis of any biomolecules and is not limited to
proteins insofar as two constituent fragments are
10 available by chemical synthesis or other biosynthetic
methods. In particular, the present process: (1)
allows rapid synthesis of modified proteins;(2) avoids
the use of protecting groups in at least the critical
bond formation stage;and (3) incorporates into the
15 protein backbone structural units such as D-amino
acids and amino acid analogs.

This invention will be described in further detail
below by way of the aforeindicated embodiments, but
20 these embodiments should not be taken as limiting the
scope of the invention.

- 21 -

EXAMPLE 1

SYNTHESIS OF [NHCH$_2$COSCH$_2$CO)$^{51\text{-}51}$Aba^{67,95}] HIV-1 PR

5   Two peptide segments, HIV-1 PR (1-50, Gly^{51}SH)
(PREPARATION 1) and bromoacetyl (53-99) HIV-1 PR
(PREPARATION 2) were coupled by ligating the segments
in 6M guanidine hydrochloride 0.1 M sodium phosphate
at pH4.3. The segments were separately dissolved in
10   ligation buffer at a concentration of 20 mg/ml.

The process of the ligation was followed by reverse
phase HPLC on a Vydac C$_{18}$ column using a linear
gradient of 30-60% buffer B (90% acetonitrile/0.09%
15   trifluoroacetic acid) in buffer A (0.1%
trifluoroacetic acid) in 30min. The flow rate was 1
ml/min and absorbance was monitored at 214 nm.
Results are shown in FIGURE 2.

20   The resulting ligated peptide was purified by reverse
phase HPLC on a semipreparative Vydac C$_{18}$ column using
various linear gradients of 90% acetonitrile/0.09% TFA
in 0.1% aq TFA. The purity of the product peptide was
checked by analytical HPLC as well as by ion spray
25   mass spectrometry. The product peptide had the
expected molecular weight of 10.768.6 ± 1.1 Da (Calcd.
for monoisotopic; 10.763.9 Da; (average) 10, 770.8
Da). The mass spectra of the title peptide are shown
in FIGURE 3A and 3B.

30

- 22 -

EXAMPLE 2

SYNTHESIS OF [(NHCH(CO NH$_2$) CH$_2$ SCH$_2$ CO]$^{51-52}$Aba^{67,95}] HIV-1 PR

Two peptide segments HIV-1 PR (1-50, Cys^{51} amide)
(PREPARATION 1) and bromoacetyl (53-99) HIV-1 PR
(PREPARATION 3) were coupled substantially according
to the procedure of EXAMPLE 1, except that the pH was
over 7.0. The ligation reaction was monitored by
HCLP, with the results shown in FIGURE 3. After work-
up and purification, the product peptide had the
expected molecular weight of 10,800.31 ± 0.75 Da
(Calcd. for monoisotopic 10,792.9 Da; (average)
10,799.8 Da).

EXAMPLE 3

ENZYMATIC ACTIVITY OF [(NHCH$_2$COSCH$_2$CO)$^{51-52}$Aba^{67,95}] HIV-1 PR

An aliquot of the ligation reaction mixture after 3
hours ligation (EXAMPLE 1) was treated with a peptide
substrate (1 mg/ml) having the sequence of Ac-Thr-Ile-
Nle-Nle-Gln-Arg. NH$_2$ (Nle:L-norleucine) at pH 6.5.
Reaction was monitored by reverse phase HPLC using a
Vydac C18 column with the results shown in FIGURE 5.
In FIGURE 5 the upper panel shows the peptide
substrate peaks before treatment. The lower panel
shows the peaks of cleavage products after 15 min.
treatment.

The cleavage products were separated by reverse phase